Voices Inside Schools

Beyond NCLB and AYP: One Superintendent's Experience of School District Reform

RON SOFO
Freedom Area School District, Pennsylvania

In this Voices Inside Schools essay, Ron Sofo, a school district superintendent in western Pennsylvania, argues that schools need bottom-up solutions more than top-down mandates if they are to prepare all students to meet twenty-first-century workforce demands. Framing the discussion in terms of his district's reform model — the "New 4 Rs" of rigor, relevance, relationships, and reflection — Sofo describes how one middle school developed a multifaceted, classroom-level intervention to support struggling learners. He then explains how that effort was scaled up to other grades and how insights from the initiative reverberated throughout the district. In telling his story, Sofo depicts the instructional reforms his staff undertook, the challenges they encountered, and the early indications of their success. His essay offers a window into the complex process of instructional reform at the classroom, school, and district levels.

Like many school districts, the Freedom Area School District in Freedom, Pennsylvania, faces the challenge of engaging all students in rigorous and relevant academic work based on twenty-first-century expectations and skills. The notion that schools must prepare all students for high levels of academic success was scarcely imagined when our nation's current K–12 system was conceived and refined in the late nineteenth and early twentieth centuries. Our public schools and the learning environments within them were not originally designed to maximize every student's opportunity for academic achievement. Schools groomed an elite few for higher education and prepared others for low-skilled employment. The system presumed — and even emphasized — the inevitability of failure for most, if not all, students at some time in their academic careers (Darling-Hammond, 1997).

Harvard Educational Review Vol. 78 No. 2 Summer 2008
Copyright © by the President and Fellows of Harvard College

However, the educational landscape has changed dramatically. The major challenge today is ensuring that all students make significant progress toward the achievement of twenty-first-century academic standards and skills. This means equipping students with sophisticated cognitive and communication abilities and preparing them for some form of higher education (Murnane & Levy, 1996). Meeting this challenge requires major rethinking in the fields of educational policy and practice (Schlechty, 1990). Such rethinking must start not with top-down mandates but with teachers and principals, who hold the key to classroom- and school-level innovation.

Since beginning this school reinvention process in Freedom five years ago, we have seen test scores rise notably. But test scores tell only part of the story. When I talk to teachers, I hear about the rigorous application of academic skills, and when I visit classrooms, I see evidence of increased teacher and student engagement in learning. Furthermore, I see evidence that these changes are reaching all students. Freedom's story illustrates how the top-down nature of the federal No Child Left Behind Act (NCLB) and its narrow student assessment tool — a single, standardized test score — does not help districts equip students with the knowledge and skills they will need to confront unpredictable, real-world situations. While NCLB has acted as a catalyst for accountability, it ignores the ways in which public school operations need to be overhauled to promote faculty empowerment. Freedom's approach to instructional improvement has evolved from a specific classroom-level set of innovations, designed to help students who were slipping through the cracks, to a district-wide initiative aimed at ensuring that all students graduate with the skills they need for success in the twenty-first century.

In this essay, I share the instructional progress our district has made and reflect on the strategies we have found to be effective. I begin by describing the ways in which educators in Freedom have found that NCLB falls short in helping students develop twenty-first-century knowledge and skills, and I go on to describe a framework called the New 4 Rs, which we adopted to help cultivate those skills. I then tell the story of how we took a classroom- and school-level intervention and scaled it up to improve teaching and learning across the district. In the process, I describe the challenges we faced, the lessons we learned along the way, and the work that remains to be done.

The Policy We Have and the Skills Our Students Need

As technology has automated routine tasks and enabled the distribution of work across great distances, success in the global economy increasingly depends on the ability to communicate across contexts, solve complex problems, and adapt quickly to new technologies (Friedman, 2006; Levy & Murnane, 2004). Schools must therefore equip students with the skills necessary to succeed in the twenty-first century, including proficiency in oral and written communication, collaboration and teamwork, citizenship and ethics, and technology

literacy (Partnership for 21st Century Skills, 2006). Unfortunately, NCLB's academic accountability tool is simply too narrow to measure students' mastery of these skills. Instead, the 2001 legislation defines appropriate student achievement as scoring above a *proficient* score on a single standardized test, and both the test and the score are chosen independently by each of the fifty states. Furthermore, the tests are not designed to measure performance in any areas other than content knowledge and skills related to math and reading. Finally, the current law does not address each student's academic growth or decline from year to year.

It is not that basic skills such as reading and mathematics no longer matter; rather, students need to demonstrate mastery of these and other skills in complex and nuanced ways. Rothstein, Wilder, and Jacobsen (2007) make this point in their historical analysis of the purposes of public education in the United States since colonial times. Their study identifies eight broad categories of objectives that have persisted throughout 300 years of educational policymaking: basic academic skills, critical thinking and problem solving, social skills and work ethic, citizenship, physical health, emotional health, the arts and literature, and preparation for skilled employment. These categories suggest that the "need to knows and dos" that were important in the past remain academically and socially relevant today. However, the contemporary expectation that *all* students will achieve a high level of proficiency in these areas is revolutionary. Rothstein, Wilder, and Jacobsen give the following example to illustrate how twenty-first-century learning in social studies places different expectations on schools: "What if schools were held accountable not for whether students could recite historical facts, but for whether they actually registered and voted?" (p. 12). The authors assert that our current accountability system under NCLB distorts school curricula with its overemphasis on basic skills. They further argue that this is happening not because of our ignorance but because of our frugality: Measuring higher-order applications of knowledge and skills costs more than we are currently spending on paper-and-pencil tests that measure basic skills. However, both the authors and I fear there is danger in this frugality. If we as educators do not challenge ourselves to think beyond the narrow mandates of NCLB, it may be the next generation that pays for our mistakes.

Beyond NCLB: The New 4 Rs

In Freedom we implemented a holistic approach to school reform that we believe will prepare our students for success in the twenty-first century. Our approach is adapted from a model developed by the International Center for Leadership in Education (ICLE) (2002). Called the "New 4 Rs" for student learning and instruction, this model moves beyond the traditional three Rs — reading, 'riting, and 'rithmetic — to emphasize *rigor, relevance, relationships,* and *reflection.*[1] Since 2003, the New 4 Rs have guided our planning and imple-

mentation efforts at Freedom, including our approaches to instruction, student achievement, professional development, new teacher hiring, and school infrastructure. We think the New 4 Rs are worthy of examination because they offer educators a path toward school improvement and serve as an example of what needs to be done to facilitate positive internal ownership of the notion of success for all students.

The first R stands for instructional *rigor*. Drawing on the ICLE framework, this goal encourages teachers to push students beyond simple cognitive functions, such as "memorizing, identifying and describing," toward "comparing, contrasting, analyzing, evaluating, synthesizing and creating," the kinds of higher-order skills students will need for success in the twenty-first century (International Center for Leadership in Education, 2002).

The second R, *relevance*, concerns how students are expected to use and apply the skills they learn. Again based on ICLE guidelines, *relevance* means challenging students to apply their learning across content areas and to demonstrate mastery by using their learning in real-world, unpredictable situations. "The performance is the understanding," a phrase borrowed from the performing arts and sports, guides what we teach, how we teach, and how we measure what students actually learn.

The third R, *relationships*, means personalizing instruction to improve student success. Our recent work teaches us that when students and staff experience meaningful human relationships that can be expressed as "we care about you and your success," both students and staff bring their best effort to the work at hand. We find that an emphasis on relationships of this kind promotes student self-confidence and responsibility for learning at high levels. Also, in order to "walk the talk" of caring about our faculty and their success, we employ an ongoing professional development program focused on high-quality content and best-practice pedagogy and driven by teachers' self-reported professional needs.

The fourth R, *reflection*, completes the cycle of continual improvement. Collaborative endeavors in fields outside of education, including athletics and the arts, remind us that the process of reflection lies at the core of creating a community of learners (Childress, 1998). If our work is important enough to apply the first three Rs of *rigor*, *relevance*, and *relationships*, then we need to find time to reflect on the impact of our efforts. Formal and informal opportunities to share lessons, plan collaboratively, and listen to students' perceptions of instruction invite both students and teachers to describe their difficulties, successes, and desires for further learning. Reflection promotes both teachers' and students' ownership of the learning process and provides the key to unlocking their best efforts to learn.

We at Freedom regard the New 4 Rs as essential elements to productive and accountable K–12 public school reform. Do we possess the collective will to undertake such reform? Harvard University professor Ronald Edmonds posed the question as follows: "We can, whenever and wherever we choose, success-

fully teach all children whose schooling is of interest to us. We already know more than we need to do that. Whether or not we do it must finally depend on how we feel about the fact we haven't so far" (Leverett, 2006).

A Reform in Freedom

Initial Challenges

The Freedom Area School District is a small district of roughly 1,700 students located about twenty-five miles northwest of Pittsburgh, Pennsylvania, in a primarily rural area of Beaver County. Our student body is ethnically homogenous but socioeconomically diverse. Less than 3 percent of our student body belongs to a racial or ethnic minority group, but approximately 37 percent of our students receive free or reduced-price meals.

I came to the district as assistant superintendent in August 1993 and was appointed superintendent in November 2002. I was fortunate to assume responsibility for a district with relative stability in its top leadership and student body. The nine-member school board had an average length of service of sixteen years per board member, and the former superintendent of schools held the position for thirty-seven consecutive years before retiring. The rate of student transience was also low, less than 10 percent overall in grades one through twelve.

The stability of my environment and my tenure as assistant superintendent meant that I had a thorough understanding of what the superintendent's role entailed. However, stability also created challenges, because change was not necessarily welcomed. Many teachers believed that most of our students were performing as well as could be expected given the district's available resources. Our NCLB state test scores usually placed Freedom in the middle- to upper-third of the fifteen county school districts and close to or somewhat above the state's averages. Yet, our high school students' performance on these high-stakes mathematics and reading exams was always lower than their eighth-grade performance had been. This decline in test scores placed our high school scores in the lower half of the county, and sometimes the state. Our high school students were making Adequate Yearly Progress (AYP) by a slim margin of only one to five percentage points. Still, the fact that we were making AYP at all was a source of reassurance for some. The danger of meeting such benchmarks is complacency. Like so many districts around the country, we were not in severe duress, but neither were we helping our students excel.

I knew the district needed to do a better job preparing our students for the global economy of the twenty-first century. *Proficient* scores on the state's NCLB-mandated tests represent entry-level capabilities needed by all of our students so they can apply their knowledge and skills to novel situations. However, on average, only half of our high school juniors earned *proficient* scores

in reading and math. If we were going to prepare our students for success beyond secondary school, improvement would be necessary.

Another challenge I faced was a limited budget. From 2000 to 2005, Freedom's average combined elementary/secondary expenditure level was $6,157 per student. This amount ranked second-lowest out of the fifteen school districts in Beaver County. Local measures of wealth, as monitored by the Pennsylvania Department of Education's ratio of real estate market value to personal income aid ratio, placed the district in the lowest third of the state's 501 school districts in total wealth. Given our low per-pupil funding rate and unfunded NCLB mandates, I recognized that our reforms had to make good use of scarce district resources.

Middle School Math Target Program Implementation

During the summer of 2003, two years after the passage of NCLB, our middle school principal, Dr. William Renko, decided that having between 25 and 47 percent of students performing at the *basic* or *below basic* level on the state's high-stakes NCLB math test was not acceptable.[2] A former researcher in curriculum and supervision and a diagnostic teacher and supervisor, Dr. Renko knew that the only way to improve student learning and test scores was to improve the quality of instruction provided by their caring teachers. Together, as principal and superintendent, we agreed to focus on math instruction and student achievement for this portion of the students entering sixth grade.

Because it involved trying a new instructional approach and measuring its effectiveness, our pilot test of the "Target Sixth-Grade Math" program (henceforth, Target Math) acted as an action research project serving twenty-three entering sixth graders in the 2003–2004 school year. The students were selected based on three criteria: They had all scored *basic* or *below basic* on the fifth-grade Pennsylvania System of School Assessment (PSSA) math test; they had never earned a grade of "D" or higher in math in grades one through five; and they were not eligible for special education services. The two teachers recruited to lead the Target Math initiative in the middle school included a sixth-grade math teacher and a math coach, both of whom were also K–6 teacher leaders/trainers in the district's Math Science Partnership initiative.

The principal and these two teachers designed the instructional model and learning environment of the Target Math program. The instructional model included eight core components intended to emphasize the New 4 Rs of *rigor, relevance, relationships,* and *reflection.* In the paragraphs that follow, I describe each of the eight components in what my ongoing reflection has shown me to be their order of importance:

1. *"A-B, Not Yet" Grading.* How do you engage students in rigorous academic learning when many of their academic learning experiences have been negative or ineffective? The Target Math team theorized that giving Fs to students who already had a history of failure in math would be counterproductive to the program's goals of improved student effort and mastery of math knowl-

edge and skills. "A-B, Not Yet" grading meant that students kept revising nonproficient work (scored "Not Yet") until its quality was high enough to warrant a true A or B. By insisting that all students perform at high levels, this strategy reinforced *rigor,* the first of the New 4 Rs. This strategy was also used to give students hope about their academic potential and to send a message that, with effort on their part, learning math at a *proficient* level was within every student's reach. In other words, success was to become the only option.

2. *Looping.* If the goal was students' mastery of sixth-grade math knowledge and skills, looping was an important element of the "glue" holding this elevated set of learning expectations together. The pilot program had one of the coteachers stay with these students over multiple years. One purpose of this strategy was to make better use of the class time teachers typically spend getting to know students' strengths and weaknesses at the beginning of the school year. By using class time more efficiently, looping aimed to reinforce the first of the New 4 Rs, *rigor.* It was also intended to reinforce a heightened level of personalization, a commitment to student success, and the building of positive relationships between students and their teachers — which hearkens back to the second of the New 4 Rs, *relationships.* The principle of looping has long been evident in extracurricular activities that foster supportive relationships between adults (e.g., sponsors, coaches, or directors) and students (e.g., debaters, athletes on teams, marching band members) over multiple years. The time gained through looping enabled the teachers to provide more occasions for practice, retesting, or helping students acquire more favorable attitudes toward the mastery of concepts and skills.

3. *Coteaching.* To support both teachers and struggling students in this new mastery learning environment, we placed two teachers with the Target Math group to provide the level of differentiated instruction and positive relationships students needed in order to experience real success. The teachers' placement was intended to foster the first two of the four Rs, *rigor* and *relationships.* Coteaching also acted as a support system for these teachers. Furthermore, coteaching set out to strengthen the fourth R, *reflection,* by increasing the frequency of teachers' dialogues about their instructional practice.

4. *Listening to students.* To improve student learning, teachers must first listen to their students and then, in the spirit of the New 4 Rs, reflect on their practice. We used the Diagnostic Instrument of Supervision (DIOS), based on Seager's (1965) Model of Instructional Supervision, to listen to students' perceptions of instruction and their classroom learning environments. The student survey contains twenty-nine items related to elements of effective teaching and learning. The purpose of the survey is to sensitize teachers to students' perceptions of three important aspects of the learning environment: teacher warmth, the level of cooperation between teachers and students, and teacher fairness. By administering the DIOS every six to nine weeks throughout the school year, teachers compared student-perception data with their own observations and

data to assess the effectiveness of their teaching strategies and the students' learning. Teachers then discussed the results of the survey with their students and used the results to generate their own reflection. According to the original Target Math teachers, this type of reflection helped them improve their practice; it also increased students' effort and sense of ownership by showing that their perceptions were heard.

5. *Guided practice.* Through the use of guided practice, students were closely monitored doing independent work within the classroom, especially as they began their homework assignments each day. Each of these students had a history of not fully completing or attempting to complete homework. On the assumption that practice not only makes perfect but also "makes permanent," teachers assigned less homework than before but gave more in-class assistance in order to build students' skills and confidence. Thus, students were encouraged to give their best effort to complete a focused set of meaningful math practice items. All students now experienced completing their homework assignments correctly before they left the classroom. This represents another example of "walking the talk" of the New 4 Rs by focusing on *rigor* and *relationships.*

6. *Computer-assisted remediation.* To extend the amount of time students spent mastering essential math concepts and skills, target students received twenty hours of computer-assisted remedial instruction beyond their regular math class. The software used included an initial assessment of math concepts and skills, which individually "placed" students in appropriate computerized teaching and practice activities. Teachers received a baseline report. The software updated this report as students progressed through their self-paced lessons. These data were also used as a check of formative progress observed in the math classroom. By emphasizing the correct diagnosis of students' academic strengths and weaknesses (with additional time for students to reach proficiency), this component strengthened student-teacher relationships and guided teachers in fostering students' mastery of core academic knowledge and skills.

7. *Modified assignments.* The pilot teachers observed that students demonstrated greater effort and mastery of complex skills when they were given fewer problems to solve. In an effort to foster *rigor,* teachers designed modified assignments to reduce the number of items presented on tests, the number of practice problems used during class, and the amount of homework assigned. They believed that the quality of student effort was more important than the quantity of work they completed. This approach was especially important for these students, who had not previously been successful in math. While some may be concerned that this practice would have the effect of narrowing the curriculum, our target faculty and middle school principal believed that having all twenty-three target students positively engaged in giving their best effort to learn the core sixth-grade math standards was essential to these stu-

dents' long-term math achievement. This goal did mean that these students did not receive instruction in certain topics (amounting to six weeks of the sixth-grade math curriculum). However, teachers judged that the omitted topics, such as Roman numerals, were peripheral to students' mastery of such essential content as decimals, fractions, and measurement.

8. *Ability grouping.* All twenty-three students who scored *basic* or *below basic* on the spring 2003 Pennsylvania Math Assessment Test were grouped for math in one common instructional period. This ability-grouping strategy was employed primarily to concentrate the teachers' efforts and — consistent with the New 4 Rs — to ensure that instructional material was relevant to students' background knowledge. During the first year of the Target Math program, one of the teachers had other instructional responsibilities throughout the school, so placing the target students together in a single class meant that she would be available to work with all of them. Furthermore, the principal believed that the concentration of all eight Target Math strategies could be most effectively monitored and adjusted if they were applied to a single group of students in one class period.

Target Program Implementation

As the 2003–2004 school year progressed, the target students showed improvement in math learning and student achievement as evidenced by classroom grades and computer-assisted math progress reports. Some students earned "Not Yet" during the grading period; however, they were willing to keep working, and all had improved their classroom grades to As and Bs by the end of the first 90-day (fall semester) grading period. Students also perceived improvement in their teachers' instructional skills. On the DIOS survey, students reported substantial improvement (of roughly 0.9 points on a 3-point scale) in three instructional areas: Recognition of Needs, Agreement on Objectives, and Assessment of Concepts and Skills. Students reported the greatest gains in their sense that "there is never any undue fooling around" in the Target Math class, and that "learning in this class is not impossible."

The 180-day (end-of-year) report cards continued to provide encouraging, positive evidence of the target program's impact on student learning. Each of the twenty-three students completed the required work for the school year and received a letter grade rather than a "Not Yet"; the lowest student grade earned during the last nine weeks was a B-. However, the class did cover less material than the other sixth-grade math classes, about two months', or forty lessons', worth. Teachers decided that because the students had a solid foundation in the material that had been covered, they would be well positioned to cover the additional material during the next school year, when teachers would make a special effort to include it in the curriculum.

Of greater concern was our finding that the positive work ethic the target students were now bringing to their math class was not transferring to their other classes. Not sure of the reason for this pattern, we wondered if it had to

do with the strong relationships students had developed with their teachers in the Target Math program. If so, it seemed essential to try to foster these relationships in other content areas.

The achievement results of the target students after twenty-two hours of work in the computer-assisted math remediation program were also encouraging. All students showed improvement after a 45-school-day pilot implementation period. In particular, seventeen students showed gains of more than one grade level; five students made gains of one and a half grade levels, and two students made gains of two grade levels.

By the end of the school year, students perceived that their teachers' instructional skills had remained constant or improved since mid-year. This is especially impressive because data from the DIOS in the late spring often trends in a significant negative direction.[3]

Our goal of keeping all middle school math and reading teachers informed of the target program's implementation efforts was a decision we made prior to implementation. Thus, throughout the first year of the Target Math program, other middle school teachers of math and reading were debriefed every nine weeks by the two Target Math teachers as a way of introducing staff to the program's strategies, successes, and challenges. This first year of observation and open discussion by the rest of the middle school math and reading faculty gave the building principal and me a set of informal needs-assessment data (particularly related to teacher values, beliefs, and instructional practices) with which to plan staff development activities that would support an expansion into all middle school math and reading classes in future years.

We were interested to learn that while the Target Math teachers reported an improvement in job satisfaction and beneficial effects on students' math learning and their own teaching, perceptions of the pilot program by other teachers in the school were not as positive. At the end of the first year, the large majority of nontarget teachers regarded the pilot program's "A-B, Not Yet" grading system as a reward to students for lack of effort in previous years and as a strategy that, according to some teachers, "poisoned other (non-target) students" (Sofo & Renko, 2005, p. 47). The Target Math teachers heard from many of their colleagues that a better way to handle or "motivate" target students was to fail them in both the elementary and middle school grades (Sofo & Renko, 2005). The presence of this perception, despite positive achievement data and DIOS data, convinced us that the next step was to engage a larger group of teachers in the program.

Moving Forward: Years Two and Three of Adaptation, Compromise, and Expansion

Given our discovery in year one that students' efforts in Target Math were not carrying over to their other classes, the middle school added a Target Reading program in year two (2004–2005). To prepare reading teachers before the school year commenced, the principal led a summer training workshop

with assistance from the Target Math teachers. As we expected from concerns raised the previous year, the reading teachers were, in particular, wary of the "A-B, Not Yet" approach to grading.

The second year also brought some unanticipated challenges, including staffing changes in the middle school that directly affected the target programs. Before year two began, the Target Math teacher who had been the school's math coach, and who was going to "loop" with the first cohort of students as they entered seventh grade, moved out of state due to her spouse's work assignment. Our original plan was that the remaining Target Math teacher (henceforth, the "main classroom teacher") would continue to work at the sixth-grade level, while the looping teacher worked with the first cohort of students as they became seventh and eighth graders. However, when the looping teacher left the system, the main classroom teacher moved up to seventh grade so the first cohort of students would have teacher continuity as planned. We filled the vacant teacher position with a veteran fifth-grade teacher who had already worked as a teacher trainer with the main classroom teacher. Consequently, the two shared solid knowledge of best practices and had a strong working relationship. These factors made the staffing transition smoother than it might otherwise have been.

Another personnel change occurred in October of year two, when Dr. Renko, the middle school principal, retired. His successor, Robert Gallagher, had worked in the district for more than twenty-eight years as an elementary reading teacher, instructional support teacher, and elementary school principal, and he was knowledgeable about the target program. However, as he stepped into the middle school role, he chose not to require a major feature of the pilot program, the "A-B, Not Yet" grading system. His concerns about adoption of this approach were numerous. First, he wondered how teachers would accurately describe what "Not Yet" students need to learn when work ranging from a traditional F to a C could garner a "Not Yet" grade. Second, he wondered whether it would be more consistent to use a grading system with the same nomenclature as the PSSA test — that is, *advanced, proficient, basic,* and *below basic.* Third, students who fail a course in grades seven through twelve are traditionally required to attend a county-operated and parent-paid summer school. Since students would not fail courses under the "A-B, Not Yet" approach, the principal wondered whether the district would have the financial and staff resources to operate its own summer school and bring each "Not Yet" student to proficiency. Finally, if the district could not afford to run its own summer school, he wondered where the extra time would come from during the school year to address the specific learning needs of students who were performing at the "Not Yet" level.

As Target Math and Target Reading were being implemented, another amendment to their core components involved the use of the DIOS student survey. Mr. Gallagher valued the use of student surveys to gather feedback on student perceptions of classroom instruction. However, teachers expressed

concern that the DIOS was time-consuming to administer and score. Moreover, some teachers saw the DIOS as a potential source of data that principals might use to compare and evaluate teachers. Mr. Gallagher had to weigh these concerns with his understanding that student feedback is very valuable in serving students more effectively. This was a difficult balance to strike. Though he was determined to give serious consideration to the continued use of the DIOS or another research-based student survey, Mr. Gallagher put the survey on hold until teachers' concerns could be addressed in some manner, perhaps through supplemental professional development.[4]

While the new middle school leadership might be perceived by some as an impediment to the change process already under way, I saw it as an opportunity to consider the target program's reform strategies through fresh eyes. Mr. Gallagher's reflections prompted us to consider teacher ownership, anticipate teachers' aforementioned concerns, and design future implementation strategies to maximize student and teacher success.

Given that it was now optional, teachers new to the target programs did not adopt the "A-B, Not Yet" grading system. The original sixth-grade teacher, who moved to seventh grade in year two, continued to use "A-B, Not Yet," but she modified it by including the grade of C, resulting in an "A-B-C, Not Yet" system, and the students in the first Target Math cohort were graded with this system through seventh grade. While the decision to discontinue use of the new grading system departed from the original pilot design, the faculty and new principal did embrace the notion that instruction needed to be significantly changed to meet the needs of target reading and math students.

The eighth-grade math teacher in year two did not teach a Target Math section. This was allowed because we all concurred that changing the classroom environment for a group of low-achieving eighth-grade math students would be difficult given students' unfamiliarity with the approach. It made sense for the eighth-grade teacher's first experience of the target program to be with the original group of sixth-grade students who had already become familiar with the program's rigor and routines. They would enter eighth grade with a "looping" teacher/instructional coach who knew them from the seventh-grade math class, and the looping teacher and eighth-grade math teacher would work together in year three to implement the target program. This decision was consistent with our goal of creating a culture of success for teachers as well as students.

Starting with year three, there were target students in reading and math in all middle school grades. In that year, however, teachers became troubled by some unintended consequences of the homogeneous grouping of all target students in one reading and one math section per grade. Pilot students entering eighth grade that year perceived that they were the "dumb" class. This perception greatly concerned the students' teachers, who requested a program modification. Teachers also reported that students' off-task behavior was more challenging to manage in a homogeneous environment. In response

to these concerns, homogenous grouping was dropped the following year in favor of spreading the low-performing students among sections. However, this alteration has made it more difficult to concentrate our limited resource of the "looping teacher/coach." When target classrooms were homogenous, one looping teacher/coach per school could address two sections per grade (one math and one reading section in grades five through eight), for a total of eight sections per day. Under heterogeneous grouping, the class periods in need of service expanded from eight to twenty-four, so we hired a teacher assistant (who is actually a certified teacher) to work under the direction of the looping teacher/coach and assist the regular classroom teachers. While the teacher assistant has provided direct instructional services to target students within the heterogeneous classroom model, we have perceived a weakening of the caring bond between the looping teacher and target students over time. This situation, and the relative merits of heterogeneous versus homogenous classrooms more generally, will need to be reexamined as we continue to refine the target program.

Another unanticipated teacher-initiated modification was the expansion of math instruction from forty-two minutes to eighty minutes per day in grades five and six. This expansion came as a result of our two pilot teachers' participation in a 2004 summer mathematics academy sponsored by the regional Math Science Partnership. Within days of their academy participation, the teachers called for a meeting to explain to the building principal and me their rationale for the additional instructional time: If all students were to understand math concepts and skills deeply enough to apply them to complex real-world problems, more class time would be needed to implement an inquiry-based instructional approach. This expansion of instructional time was also adapted by the eighth-grade reading and language arts teachers; they combined the 42-minute periods for each subject into one "blocked" period in which the two subjects could be taught as a set of interdisciplinary skills.

These challenges we encountered in years two and three forced us to consider multiple perspectives and adapt to changing circumstances — steps any reform must take if it is going to survive. As with any effort to expand a small, innovative program beyond its initial setting, there were growing pains. But there was also the excitement of building on a homegrown program that seemed to reach struggling learners. The tension lay in determining how far the program could be adapted without sacrificing the features that made it powerful. This is a tension that the middle school will continue to address. It is also a tension we face at the district level as we consider how to apply lessons learned from the middle school reform more broadly.

Looking Toward High School Reform

In 2004–2005, as the middle school target program evolved, our high school also embarked on a reform journey. Under the leadership of its principal, Dr. Robert W. Staub, our good twentieth-century comprehensive high school

began working toward becoming a great twenty-first-century high school that would meet the needs of all students. The high school initiative seemed essential if the district, as an organization, was to take advantage of what we were learning through the implementation of the middle school Target Math and Target Reading programs.

Through a state-funded high school reform grant called Project 720 (named for the number of instructional days in four typical school years), Dr. Staub and I engaged the faculty, community, and school board in a process of reflection, data collection, goal setting, and planning. Our mission was to chart a four-year course for transforming high school teaching and learning so that all students could master rigorous and relevant skills for the twenty-first century. Guided by ICLE (2002) research on effective high schools and by an ICLE consultant, Dr. Gary Fields, we wrote a report identifying our high school's current strengths and its opportunities for improvement. In an effort to foster the New 4 Rs at the high school level, we then took members of our faculty and the Project 720 steering committee (comprising members of the faculty, administration, school board, and community) to see secondary school programs around the country that ICLE had identified as using "best practices." In particular, our inquiry focused on rigor and relevance. Teachers involved in these early efforts then led staff development experiences to help their colleagues understand and deliver rigorous and relevant lessons using project-based activities. Our belief was that a supportive program of staff development would provide a strong foundation on which to promote academic success for all students.

As with the middle school target programs, we worked to foster relationships at the high school level by listening to teachers' needs. In some cases, those needs were strikingly similar to what we heard at the middle school level: Working independently from their middle school counterparts, high school faculty determined that the eight-period, forty-two-minute-per-period school day could not adequately support rigorous, relevant instruction facilitated by technology-enhanced, project-based learning. As a result, the high school proposed — and the school board adopted — a modified block schedule for the 2007–2008 school year. Clearly, this teacher-led recommendation for expanded high school instructional time paralleled a change in the middle school, and for essentially the same reasons. That these changes were made independently of one another attests to teachers' increased attention throughout the district to research on best practices.

Moreover, our experience pilot-testing new ideas in the middle school informed our approach in the high school. In the year prior to official block-schedule implementation, we piloted seven days of block-schedule instruction. The test allowed both faculty and students to gain experience with the approach and to try using the innovative lessons it permitted. We then collected data from student focus groups, faculty surveys, and in-service discussion groups to inform our implementation process during the new school

year. Moreover, we gathered feedback on how to differentiate staff development based on teacher comfort, performance, and need.

Also similar to the middle school reform, a major feature of the high school's Project 720 plan was a zero-tolerance policy for student failure in ninth grade. The steering committee had identified a pyramid of progressively intensive interventions that teachers and counselors could use to help freshmen avoid failing their courses. However, when members of the steering committee briefed me about the genesis and components of the ninth-grade success model, I initially questioned the need for this portion of the reform. Given the skepticism with which middle school teachers had met a success-only policy, I wanted the committee to understand what they would be taking on. In order to gauge teachers' and administrators' commitment to the idea — and their possible willingness to consider the "A, B, Not Yet" grading approach — I deliberately challenged the rationale behind zero tolerance for ninth-grade student failure. I reminded the committee that this grading approach embodied a concrete, everyday, visible practice that would require everyone to "walk the talk" of supporting all students' academic success. However, I also wanted the committee to understand the potential benefits of the "Not Yet" approach. I took the opportunity to share the positive preliminary student achievement data on the impact of the first-year Target Math program, which had the "A-B, Not Yet" grading system as one of its elements.

The steering committee representatives emphasized how important this portion of their plan was to the long-term success of the high school reform. They indicated the need to more closely coordinate with our middle school faculty and seventh- and eighth-grade students to help prepare them for high school success before the freshman year. I then asked if the high school would consider adopting the "A-B, Not Yet" grading approach for ninth grade. I also suggested that possibly a C grade might be appropriate to use at the high school level.

My question was met with silence, which was broken only when one committee member asked, "What would we do with students who perform at the 'Not Yet' level?" I said that the pyramid of interventions would be employed. "But what happens if we have students who are 'Not Yet' at the end of the freshman year?" I suggested that the interventions needed to be used starting the first week of school. If any interventions did not appear to be effective based on the data the school collected, the staff would need to modify or drop those interventions from the pyramid. For students who were still "Not Yet" by the end of the school year, we would tailor a detailed set of curriculum standards specifically to their needs, and we would provide a focused summer school or afterschool program to help them meet those standards.

After further discussion and reflection, Dr. Staub decided it would be appropriate for one or two of the ninth-grade teachers to pilot the "A-B-C, Not Yet" grading approach in the next school year. His reluctance for the entire ninth-grade team to implement the new system stemmed from the great number of

instructional initiatives scheduled to be carried out during the year, including project-/problem-based learning, wireless student laptop computing, and SMART Board instruction, as well as a differentiated approach to staff development that meets teachers' individual learning needs. The ninth-grade teachers chosen to pilot the program were those whose practice already embodied the core principles of the "Not Yet" approach. Dr. Staub believes that a mini-pilot program can showcase the positive impact of the new grading system on engagement and performance and offer some initial success stories that pilot teachers can share with their peers.

Both my exchange with the committee and Dr. Staub's subsequent reflection on a success-only policy represent the systemic approach and quality of thinking that the mandates of NCLB fail to recognize and foster. NCLB assumes that educators must work harder within the current K–12 education system to enable all students to achieve academic success. Yet, expecting educators to work more effectively within the current system to "leave no child behind" is like asking potential astronauts to use a Model T Ford to carry passengers to the moon. The possibility of reaching the moon in a Model T or fulfilling the vision of NCLB is not only affected by the talent and skills of available personnel but also by the tools — including systems — with which they have to work.

Preliminary Results

While we cannot claim definitively that our reforms are the sole source of student achievement gains in Freedom, early results are encouraging. Of the seventeen students who remained in the pilot Target Math program for grades six, seven, and eight, 59 percent (ten students) earned scores of *proficient* or *advanced* on the eighth-grade PSSA mathematics test. Middle school teachers and administrators who knew these students in fifth grade predicted that if they had been placed in traditional mathematics classrooms, no more than two of them would have performed at the *proficient* level on this high-stakes test. The remaining seven students each faced extraordinary family and life circumstances outside of school that made giving their best effort every day a tremendous challenge.

On a larger scale, the proportion of 2003 fifth graders who earned *proficient* or *advanced* scores in mathematics rose from 53 to 79 percent by 2006 (the students' eighth-grade year), for a gain of 26 percentage points. By way of comparison, the state's percentage of students scoring *proficient* or *advanced* rose only six points in the cohort of 2003 fifth graders over the same time period. In other words, Freedom students' increase in the percentage scoring *proficient* or *advanced* from fifth to eighth grade was over four times the state average. Furthermore, among students tested in 2003 or 2004 in grades three, five, or eight, the percentage scoring *proficient* or *advanced* in 2006 increased in all student subgroups. This improvement in a cohort's achievement scores

from eighth to eleventh grade is a first for our district. Moreover, test scores improved again in 2006–2007, which suggests that our progress the previous year was not merely a fluke but part of an upward trend.

These results include at least 90 percent of the same students taking the test from the baseline year to 2006. Again, I cannot say that these achievement gains are due entirely to our reform efforts, but this is the first time Freedom has seen growth in all three cohorts since these tests were initially administered twelve years ago. Bearing in mind the limits of external accountability measures, we take this sustained growth as an encouraging sign that we may be on the right track.

Reflection

Is our nation willing to support a system of public education that places students' mastery of twenty-first-century knowledge and skills at the top of the priority list? I hope so. The truth is that we do not have a choice if we want to remain competitive in the global marketplace. However, ensuring that all students develop these skills will require rethinking the core assumptions that drive most current educational policy and practice. In Freedom we are headed toward creating a new, responsible, effort-based, success-driven system of education that goes beyond the legal mandates of NCLB and touches the hearts and minds of our students and staff — a system based on *rigor, relevance, relationships,* and *reflection.* However, we still have much more to do.

The district needs to continue its work to build a culture of student and faculty success. Most of our teachers, while supporting the concept of student success, find it difficult to adopt a grading system that does not have failing grades as part of the rating system. Our current middle school faculty has articulated a series of questions that have not been answered sufficiently for them to embrace the "A-B, Not Yet" approach to student grading and motivation. These questions include the following:

- What do we do for the students who are still performing at the "Not Yet" level at the end of the school year?
- How would this affect their transcripts?
- Is it fair that these students receive a higher Quality Point Average than students who are succeeding and working hard in classes where their teachers do not use this grading approach?
- Is it fair that some students make honor role with the "A-B, Not Yet" grading system and other students do not because they receive a "C" grade in math class?

Still, as superintendent, I continue to wonder whether our current system of grading has more to do with control and power within the student-teacher relationship than with maximizing each student's potential for learning and academic growth.

Finding adequate time for professional development within a 187-day teacher contract year (which includes only 180 days of student instruction) makes the work of "retooling" quite challenging, much like trying to rebuild an aircraft in flight while it is full of passengers. Working collaboratively to redefine a professional work year for all teachers will take time. But as more teachers set goals that are difficult to attain within the constraints of the academic calendar, we may see teachers willing to consider a longer work year that incorporates more time for professional learning.

Also, having most teachers embrace looping, in which they stay with the same group of students for more than one year, is a slow process, especially at the middle and high school levels. However, students in the upper grades have long known that they could keep the same teacher for multiple years by taking electives in which that teacher specialized. This informal type of looping has existed for years, but its availability has depended on students' grade level and course of study. If we want to build a system of public education that cares for all students equally, it may be worthwhile to make looping more widely and formally available in the middle and upper grades.

Despite these challenges, preliminary results indicate we are moving in the right direction. Author and futurist John Schaar has famously observed that "the future is not some place we are going to, but one we are creating" (Joseph Rowntree Charitable Trust, n.d.). In the 1960s, our nation set the goal of placing a man on the moon. We reached that goal by finding ways of integrating existing knowledge with new information and applying innovative technologies in unconventional and courageous ways. Creating a K–12 system of educational excellence and success for all students and teachers will require a rekindling of that spirit that took us to the moon. Our quest for excellence for all students and their teachers requires us to go beyond external bureaucratic models of control and compliance. We need to transform ourselves at the local level into organizations that can sustain successful learning environments based on trust, disciplined dialogue, shared values, and data-driven decisionmaking (Schlechty, 2006).

Freedom's goal is to keep moving in that direction. I have come to regard strengthening individual relationships within the district as an important key to our success. Our former middle school principal, William Renko, observed that "nothing matters in education — money, legislation, big names, big business, Harvard — if teachers don't connect with kids." Central office administrators can help facilitate those connections by ensuring that they actively listen to and connect with teachers and principals. This active listening needs to be characterized by treating staff as equal partners rather than subordinates. If we are to prepare students to succeed in the unpredictable world of the twenty-first century, our relationships with faculty and fellow administrators should model the supportive relationships that we want teachers to establish with students. Built on this foundation, success for all students and teachers will become the only option.

Notes

1. The New 4 Rs share their origins within the eight components of the 2003–2004 Freedom Middle School's Target Math Program and ICLE's (2002) rigor and relevance model of instruction. This ICLE model was initially adopted by our high school and now has been adopted by the entire district.
2. Pennsylvania's State Board of Education recognizes four levels of performance on the state assessment: *advanced, proficient, basic,* and *below basic.* For purposes of NCLB's measurement of AYP, *advanced* and *proficient* performances are considered to be "above the bar," while *basic* and *below basic* are "below the bar."
3. This observation by Dr. Seager is based on data collected by a number of doctoral students at the University of Pittsburgh and reported in their unpublished dissertations, of which Dr. Renko's was one.
4. More recently, the middle school has moved away from ability grouping and has begun grouping target students and nontarget students heterogeneously. While this measure is intended to reduce social stigma for target students, it means that collecting data specific to these students has become more difficult. In light of this data-collection obstacle, we need to revisit how to implement the DIOS in future years. This may mean providing more staff development to middle school teachers on the use of student surveys.

References

Childress, H. (1998). Seventeen reasons why football is better than high school. *Phi Delta Kappan, 79,* 616–619. Retrieved February 11, 2008, from http://www.pdkintl.org/kappan/kchi9804.htm

Darling-Hammond, L. (1997). *The right to learn: A blueprint for creating schools that work.* San Francisco: Jossey-Bass.

Friedman, T. L. (2006). *The world is flat: A brief history of the twenty-first century.* New York: Farrar, Straus, & Giroux.

International Center for Leadership in Education. (2002). *Rigor and relevance handbook.* Rexford, NY: Author.

Joseph Rowntree Charitable Trust. (n.d.). *Visionary quotations.* Retrieved February 11, 2008, from http://www.jrct-visionaries.org.uk/page.asp?section=000100020008

Leverett, L. (2006, September 6). Closing the achievement gap. *Edutopia News* (E-mail bulletin).

Levy, F., & Murnane, R. J. (2004). *The new division of labor: How computers are creating the next job market.* New York: Russell Sage Foundation.

Murnane, R. J., & Levy, F. (1996). *Teaching the new basic skills.* New York: Free Press.

Partnership for 21st Century Skills. (2006). *Results that matter: 21st century skills and high school reform.* Tucson, AZ: Author.

Rothstein, R., Wilder, T., & Jacobsen, R. (2007). Balance in the balance. *Educational Leadership, 64*(8), 9–14.

Schlechty, P. C. (1990). *School for the 21st century: Leadership imperatives for educational reform.* San Francisco: Jossey-Bass.

Schlechty, P. C. (2006, October). Bureaucracies and learning organizations. *The School Administrator, 63,* 62.

Seager, G. B. (1965). *The development of a diagnostic instrument of supervision.* Unpublished doctoral dissertation, Harvard Graduate School of Education, Cambridge, MA.

Sofo, R., & Renko, W. (2005). *No bad schools: On the high road to educational reform.* Bangor, ME: Booklocker Digital Press.